

A Note on Ordinal DFAs

S.L. Bloom

Department of Computer Science
Stevens Institute of Technology
Hoboken, NJ, USA

YiDi Zhang

Department of Mathematics
Stevens Institute of Technology
Hoboken, NJ, USA

Abstract

We prove the following theorem. Suppose that M is a trim DFA on the Boolean alphabet $0, 1$. The language $\mathcal{L}(M)$ is well-ordered by the lexicographic order $<_\ell$ iff whenever the non sink states $q, q.0$ are in the same strong component, then $q.1$ is a sink. It is easy to see that this property is sufficient. In order to show the necessity, we analyze the behavior of a $<_\ell$ -descending sequence of words. This property is used to obtain a polynomial time algorithm to determine, given a DFA M , whether $\mathcal{L}(M)$ is well-ordered by the lexicographic order.

Last, we apply an argument in [BE10, BE10a] to give a proof that the least nonregular ordinal is ω^ω .

1 Introduction

A regular linear ordering is a component of the initial solution (in the category LO of linear orderings, see below) of a finite system of fixed point equations of the form

$$X_i = t_i, \quad i = 1, \dots, n,$$

where each t_i is a term built from the variables X_1, \dots, X_n using the constant symbol $\mathbf{1}$, denoting the one point order, and the binary function symbol $+$, for ordered sum. For example, the initial solution

$$X = \mathbf{1} + X$$

is the nonnegative integers, ordered as usual, and the initial solution of

$$X = X + \mathbf{1} + X$$

is the rationals, ordered as usual. (It is known that such systems have initial solutions in LO [BE10, Ada74, Wand79].) When the ordering is well-founded, it is a regular well-ordering.

Any countable (regular) linear ordering is isomorphic to a (regular) subset of words ordered lexicographically. (See [Cour78, Cour83].) We consider the question: which trim deterministic finite automata, or DFAs, M have the property that the language $\mathcal{L}(M)$, ordered lexicographically, is well-ordered? (Such DFAs are the “ordinal DFAs” of the title.)

As a consequence, we obtain a polynomial time algorithm to determine, given a DFA M , whether $(\mathcal{L}(M), <_\ell)$ is well-ordered.

If α is a regular ordinal, there is a trim DFA M such that $(\mathcal{L}(M), <_\ell)$ has order-type α . We obtain the known result that α is less than ω^ω by adopting to DFAs the technique in [BE10] which was applied to context-free grammars.

2 Preliminaries

We review some well-known concepts to establish our terminology. A linearly ordered set $(L, <)$ is a set equipped with a strict linear ordering, i.e., a transitive, irreflexive relation such that for $x, y \in L$, exactly one of $x = y$, $x < y$, $y < x$ holds. Here, we will assume that any linearly ordered set is at most countable. A **morphism** $\varphi : (L, <_1) \rightarrow (L', <_2)$ of linearly ordered sets is a function that preserves the ordering: if $x <_1 y$ then $\varphi(x) <_2 \varphi(y)$, and thus φ is injective. Thus, the linearly ordered sets form a category **LO**. Two linearly ordered sets are isomorphic if they are isomorphic in this category. A linearly ordered set $(L, <)$ is **well-ordered** if every nonempty subset of L has a least element. The **order-type** $\mathbf{o}(L, <)$ of a linearly ordered set is the isomorphism class of $(L, <)$. A (countable) **ordinal** is the order-type of a well-ordered set.

If $(L, <_1)$ and $(L', <_2)$ are linearly ordered sets, the ordered sum

$$(L, <_1) + (L', <_2)$$

is the linearly ordered set obtained by defining all points in L to be less than all points in L' , and otherwise keeping the original orders. More generally, if for each $n \geq 0$, $(L_n, <_n)$ is a linearly ordered set, then the **ordered sum**

$$(L_0, <_0) + (L_1, <_1) + \dots$$

is the set $\bigcup_n L_n \times \{n\}$ ordered as follows:

$$(x, i) < (y, j) \iff i < j \text{ or } i = j \text{ and } x <_i y.$$

If a set Σ is linearly ordered, the **lexicographic order** on the set of words on Σ , Σ^* , is defined for $u, v \in \Sigma^*$ by

$$u \leq_\ell v \iff u \leq_p v \text{ or } u <_s v,$$

where \leq_p is the **prefix order** and $<_s$ is the **strict order**:

$$\begin{aligned} u \leq_p v &\iff v = wu, \text{ for some } w \in \Sigma^*, \text{ and} \\ u <_s v &\iff u = x\sigma_1 w \text{ and } v = x\sigma_2 w', \text{ for some } x, w, w' \in \Sigma^* \text{ and} \\ &\quad \sigma_1 < \sigma_2 \text{ in } \Sigma. \end{aligned}$$

We write $u <_\ell v$ if $u \neq v$ and $u \leq_\ell v$. If u is the word $b_0 b_1 \dots b_{k-1}$ whose length $|u|$ is k and if $0 \leq i \leq j < k$, we write $u[i \dots j]$ for the subword $b_i \dots b_j$ of u . Also, we write $(u)_i$ for the i -th letter b_i of u . In particular, $(u)_i = u[i \dots i]$.

The next Proposition recalls some elementary facts.

Proposition 2.1 *1. For any two distinct words u, v with $|u| \leq |v|$, either $u \leq_p v$, or $u <_s v$ or $v <_s u$.*

2. If $u <_\ell v$, then $wu <_s wv$, for any word w , and conversely, if $wu <_\ell wv$, then $u <_\ell v$.

3. If $u <_s v$, then $uw <_s vw'$, for any words w, w' .

4. $u <_s v$ iff there is some i such that $u[0 \dots i-1] = v[0 \dots i-1]$ and $u[0 \dots i] <_s v[0 \dots i]$.

\mathbb{B} is the two element set $\mathbb{B} = \{0, 1\}$ ordered as usual. The set of words on \mathbb{B} , ordered lexicographically, has the following universal property.

Proposition 2.2 *For any countable linear ordering $(L, <)$ there is a subset P of \mathbb{B}^* such that $(L, <)$ is isomorphic to $(P, <_\ell)$.*

Proof. Any countable linear ordering is isomorphic to a subset of the rationals ordered as usual. But the rationals are isomorphic to the set of words on the ordered alphabet $0 < 1 < 2$ denoted by the regular expression $(0 + 2)^*1$, since this set has no first or last element, and between any two words is a third. But the ordered set $0 < 1 < 2$ is isomorphic to $0 <_\ell 10 <_\ell 11$. Thus, any countable linear ordering can be embedded in $((0 + 11)^*10, <_\ell)$. \square

A linearly ordered set $(L, <_\ell)$ is *not* well-ordered if and only if there is a sequence $(w_n)_{n \geq 0}$ of words in L such that $w_{n+1} <_\ell w_n$, for all n . In fact, sets of words that are not well-ordered by $<_\ell$ are characterized by the following lemma.

Lemma 2.3 *If $L \subseteq \{0, 1\}^*$ and $(L, <_\ell)$ is not well-ordered, then there is an infinite sequence $(w_n)_{n \geq 0}$ of words in L such that*

$$w_{n+1} <_s w_n,$$

for all $n \geq 0$.

Proof. Suppose that $(v_n)_{n \geq 0}$ is a countable $<_\ell$ -descending chain of words in L . Then, for each n , either $v_{n+1} <_p v_n$ or $v_{n+1} <_s v_n$. Define $u_1 = v_1$. Since v_1 has only finitely many prefixes, there is a least integer k such that $v_{k+1} <_s v_k <_p \dots <_p v_1$. Then define $u_2 = v_{k+1} <_s u_1$. Similarly, assuming that u_m has been defined as $v_{m'}$, for some m' , we may define u_{m+1} as the first v_k such that $k > m'$ and $v_k <_s u_m$. \square

A **deterministic finite automaton** M , DFA for short, consists of a finite set Q , the “states”, an element $s \in Q$, the “start state”, a finite set Σ , the “alphabet”, a function $\delta : Q \times \Sigma \rightarrow Q$, the “transition function”, and a subset F of Q , the “final states”. The transition function is extended to a function $Q \times \Sigma^* \rightarrow Q$ in the standard way:

$$\begin{aligned} \delta(q, \epsilon) &:= q, \quad q \in Q \\ \delta(q, \sigma u) &:= \delta(\delta(q, \sigma), u), \quad q \in Q, \sigma \in \Sigma, u \in \Sigma^* \end{aligned}$$

where ϵ is the empty word. For $q \in Q$, $u \in \Sigma^*$, we write $q.u$ instead of $\delta(q, u)$. For any state q , the **language determined by q** , $\mathcal{L}(q)$, is the set

$$\mathcal{L}(q) := \{u \in \Sigma^* : q.u \in F\}.$$

The **language determined by M** , $\mathcal{L}(M)$, is the language determined by the start state $\mathcal{L}(s)$. We say that a DFA is **trim** if for every state q , there is some word u such that $s.u = q$, and, there is at most one state q such that $\mathcal{L}(q) = \emptyset$. We call a state q such that $\mathcal{L}(q) = \emptyset$ a **sink state**.

In view of Proposition 2.2, from now on we assume that the alphabet of all DFAs is $\mathbb{B} = \{0, 1\}$.

The underlying **labeled directed graph**, $G(M)$, of a DFA M has as vertices the states of M ; there is an edge $q \rightarrow q'$ labeled b if and only if $q.b = q'$, for some $b \in \mathbb{B}$. A **strong component** of M is a strong component of $G(M)$. Recall that two states q, q' are in the same strong component iff there are paths in $G(M)$ from q to q' and from q' to q . A strong component c is **nontrivial** if there is at least one edge $q \rightarrow q'$, where both q, q' belong to c . An edge $q \rightarrow q'$ is an **exit edge** of a strong component c if q belongs to c and q' does not.

Definition 2.4 An **ordinal DFA** is a trim DFA M such that $(\mathcal{L}(M), <_\ell)$ is well-ordered.

2.1 The characterization theorem

Lemma 2.5 Suppose M is an ordinal DFA. For every state q of M , $(\mathcal{L}(q), <_\ell)$ is well-ordered.

Proof. Suppose that $(w_n)_{n \geq 0}$ is a descending sequence of words in $(\mathcal{L}(q), <_\ell)$. Since q is accessible, there is a word v such that $s.v = q$. Then (vw_n) is a descending sequence in $(\mathcal{L}(M), <_\ell)$, a contradiction. \square

The next lemma gives a necessary condition that M is an ordinal DFA.

Lemma 2.6 (Main Lemma) *Let M be an ordinal DFA. For any non sink state q , if q and $q.0$ are in the same strong component, then $q.1$ is a sink.*

Proof. Suppose, in order to obtain a contradiction, that v is a word such that $q.1v \in F$. Let u be a word such that $(q.0).u = q$. For $n \geq 0$, define

$$w_n := (0u)^n 1v.$$

Then $w_{n+1} <_s w_n$, for each n , and w_n is in $\mathcal{L}(q)$, contradicting Lemma 2.5. This contradiction shows $\mathcal{L}(q.1) = \emptyset$. \square

In any DFA, a **recursive state** q is a non sink state such that

$$q.u = q,$$

for some nonempty word u .

Now we prove the converse to the Main Lemma 2.6.

Suppose that $(\mathcal{L}(M), <_\ell)$ is not well-ordered. Let

$$\dots <_s w_{n+1} <_s w_n <_s \dots <_s w_1$$

be an infinite descending chain in $\mathcal{L}(M)$.

Say **position i is active at time n** if

$$\begin{aligned} w_n[0 \dots i-1] &= w_{n+1}[0 \dots i-1] \\ (w_n)_i &= 1 \\ (w_{n+1})_i &= 0. \end{aligned}$$

Remark 2.7 *The terminology “time” is suggested by the picture that at the n -th click of a clock, the two words w_n, w_{n+1} are generated, yielding the active position accounting for the fact that $w_{n+1} <_s w_n$.*

Proposition 2.8 *There is no upper bound on the active positions.*

Proof. Suppose otherwise. Let n be a positive integer such that all active positions are less than n . Then, by part 4 of Proposition 2.1, there would be an infinite descending sequence of words of length at most n . \square

Let i_0 be the least position which is active at any time.

Proposition 2.9 *Position i_0 is active at exactly one time t_0 .*

Proof. Suppose, in order to obtain a contradiction, that t_0 is the least time when position i_0 is active, and that $n > t_0$ is the least time after that when position i_0 is active. But then $w_{t_0+1}[0 \dots i_0 - 1] = w_n[0 \dots i_0 - 1]$ and $(w_{t_0+1})_{i_0} = 0$ while $(w_n)_{i_0} = 1$, showing $w_{t_0+1} <_s w_n$, an impossibility. \square

Corollary 2.10 *For all $n > t_0$,*

$$w_n[0 \dots i_0] = w_{t_0+1}[0 \dots i_0]. \quad \square$$

By considering the descending sequences (w_n) , $n > t_0$, we obtain the following fact.

Proposition 2.11 *There is a least position $i_1 > i_0$ which is active at a unique time $t > t_0$.* \square

In fact, the same argument proves the following.

Proposition 2.12 *There is a unique sequence $(i_k)_k$ of positions and a sequence $(t_k)_k$ of times such that for each $k \geq 0$,*

1. i_0 is the least position active at any time;
2. t_0 is the unique time when i_0 is active;
3. i_{k+1} is the least position larger than i_k active at any time larger than t_k ;
4. t_{k+1} is the unique time larger than t_k such that position i_{k+1} is active.
5. For each $k \geq 0$, if $n > t_k$, then

$$w_{t_k}[0 \dots i_k - 1] = w_n[0 \dots i_k - 1]. \quad (1)$$

\square

Example 1. Consider the sequences

$$\begin{aligned} w_1 &= 11 \\ w_2 &= 10 \\ w_3 &= 01 \\ w_4 &= 00 \\ w_k &= 00 \dots, \quad k > 4. \end{aligned}$$

Here,

$$i_0 = 0, \quad t_0 = 2, \quad i_1 = 1, \quad t_1 = 3.$$

Then position 0 is active at time 2 and position 1 is active at times 1 and 3.

Example 2. For any words u, v, w , consider the sequences

$$\begin{aligned}
 w_1 &= w1u1v1 \\
 w_2 &= w1u1v0 \\
 w_3 &= w1u0v1 \\
 w_4 &= w1u0v0 \\
 w_5 &= w0u1v1 \\
 w_6 &= w0u1v0 \\
 w_7 &= w0u0v1 \\
 w_k &= w0u0v1 \dots, \quad k > 7.
 \end{aligned}$$

Say $|w| = p$, $|u| = n$ and $|v| = m$. Then

$$\begin{aligned}
 i_0 &= p + 1 \\
 t_0 &= 4 \\
 i_1 &= p + 1 + n + 1 \\
 t_1 &= 6.
 \end{aligned}$$

From this list of words, we cannot determine i_2 , even though position $p + n + m + 2$ is active at times 1, 3, 5.

We are now able to prove the converse of the Main Lemma.

Proposition 2.13 *If $(\mathcal{L}(M), <_\ell)$ is not well-ordered, there is a recursive state q in the same strong component as $q.0$ and $q.1$ is not a sink.*

Proof. Suppose that $(w_n)_n$ is a descending sequence in $\mathcal{L}(M)$. We use the notation of Proposition 2.12. Define the state q_k by

$$q_k := s.w_{t_k}[0 \dots i_k - 1],$$

where s is the start state. By the pigeonhole principle, there are positive integers k, p with $q_k = q_{k+p}$. Then

$$w_{t_k}[0 \dots i_k - 1] = w_{t_{k+p}}[0 \dots i_k - 1]$$

by (1, part 4), so that

$$q_k = q_{k+p} = q_k.w_{t_{k+p}}[i_k \dots i_{k+p} - 1].$$

But $(w_{t_{k+p}})_{i_k} = 0$, since position i_k is active at time t_k , showing that $(w_{t_{k+1}})_{i_k} = 0$ and position i_k cannot be active after time t_k . Thus, q_k and $q_k.0$ are in the same strong component. But $(w_{t_k})_{i_k} = 1$, again, since position i_k is active at time t_k , so that

$$q_k.1 = s.w_{t_k}[0 \dots i_k],$$

which is not a sink, since $s.w_{t_k} \in F$. □

Corollary 2.14 *If M is a trim DFA, then $\mathcal{L}(M), <_\ell$ is not well-ordered if and only if there is a recursive state q in the same strong component as $q.0$ and $q.1$ is not a sink.* \square

Proposition 2.15 *Given a trim DFA M with n states, there is an $O(n^2)$ -time algorithm to determine whether $(\mathcal{L}(M), <_\ell)$ is well-ordered.*

Proof. Assume M has n states. There is a linear time algorithm, say depth-first search, to check, given states q, q' , whether there is a nonempty word u with $q.u = q'$. (see e.g., [CLRS], Chapter 22.) Then for all states q such that there is a nonempty word $q.v = q$, check to see that when there is a word u with $(q.0).u = q$, then $q.1$ is a sink. This is an $O(n^2)$ -time algorithm. \square

3 Upper bound

In [Heil80] it was shown that all nonzero regular well-orderings can be built from **1** using the operations of sum and the function $\alpha \mapsto \alpha \times \omega$. (In [BC01], these operations on words are axiomatized.) It follows immediately that the least ordinal which is not regular is ω^ω . Another method to obtain this result uses the equivalence between regular and automatic ordinals [Del04]. This note presents another argument, based on the techniques in [BE10].

3.1 Ordinals

We make some observations on ordinals.

Lemma 3.1 *The least class \mathcal{C} of ordinals containing $0, 1$ satisfying the two conditions*

- *if $\alpha, \beta \in \mathcal{C}$, then $\alpha + \beta \in \mathcal{C}$;*
- *if $\alpha \in \mathcal{C}$, then $\alpha \times \omega \in \mathcal{C}$*

is $\{\alpha : \alpha < \omega^\omega\}$. \square

We will use Lemma 3.1 to show every ordinal less than ω^ω is the order-type of $(\mathcal{L}(M), <_\ell)$, for some DFA M .

3.2 DFAs and ordinals

Definition 3.2 *Let FA be the class of ordinals representable as the order-type of $(\mathcal{L}(M), <_\ell)$, for an ordinal DFA M .*

We show that FA has the properties of Lemma 3.1.

Lemma 3.3 • *0,1 belong to FA.*

- *If $\alpha, \beta \in \text{FA}$, then $\alpha + \beta \in \text{FA}$*
- *If $\alpha \in \text{FA}$, then $\alpha \times \omega \in \text{FA}$.*

Proof. We prove only the third statement. Suppose that M is a DFA with start state q_1 . Let M' be the DFA obtained by adding a new start state q_0 to M with the transitions

$$\begin{aligned} q_0 \cdot 1 &= q_0 \\ q_0 \cdot 0 &= q_1. \end{aligned}$$

Otherwise, the states, transitions and final states are those of M . Then the set of words recognized by M' are all those of the form

$$1^n 0u, \quad u \in \mathcal{L}(M), \quad n \geq 0.$$

Thus, if the order-type of $(\mathcal{L}(M), <_\ell)$ is α , the order-type of $\mathcal{L}(M')$ is

$$\alpha + \alpha + \dots = \alpha \times \omega. \quad \square$$

Corollary 3.4 *Every ordinal α less than ω^ω is the order-type of $(\mathcal{L}(M), <_\ell)$, for some ordinal DFA M .* \square

In the remainder of this section we will prove the converse of Corollary 3.4: if α is the order-type of $(\mathcal{L}(M), <_\ell)$, then $\alpha < \omega^\omega$.

One implication of the Main Lemma 2.6 is the following.

Proposition 3.5 *Suppose that M is an ordinal DFA and q is a recursive state. Let $u_0 = u_0^q$ be a shortest nonempty word such that $q.u_0 = q$. Then, if v is any word such that $q.v = q$, then v is some power of u_0 , i.e.,*

$$v = u_0^n,$$

for some nonnegative integer n .

Proof. Suppose that $n \geq 0$ is least such that u_0^{n+1} is not a prefix of v . Write

$$v = u_0^n u x w$$

where u is a prefix of u_0 , $x \in \mathbb{B}$, and ux is not a prefix of u_0 . If $x = 0$, then $u1$ is a prefix of u_0 , since u is a proper prefix of u_0 . Similarly, if $x = 1$, $u0$ is a prefix of u_0 . In either case, $q.u$, $q.u0$ and $q.u1$ are in the same strong component, contradicting the Main Lemma. \square

We will write just u_0 rather than u_0^q when the state q is understood.

Corollary 3.6 *Suppose that M is an ordinal DFA and q is a recursive state in M . Then $w \in \mathcal{L}(q)$ if and only if for some $n \geq 0$,*

$$w = u_0^n p,$$

for some prefix $p <_p u_0$ of u_0 which belongs to $\mathcal{L}(q)$, or

$$w = u_0^n u_0 v,$$

for some words u, v such that $u1 \leq_p u_0$ and $v \in \mathcal{L}(q.u_0)$.

Proof. It is clear that any word of the above two kinds belongs to $\mathcal{L}(q)$.

Conversely, if the path starting at q determined by the word w does not leave the loop labeled u_0 , then $w = u_0^n p$, for some $n \geq 0$ and some prefix p of u_0 such that $p \in \mathcal{L}(q)$. Otherwise, this path leaves the loop after n repetitions via an exit edge labeled 0, by the Main Lemma. In this case, $w = u_0^n u_0 v$, where $u1 \leq_p u_0$ and $v \in \mathcal{L}(q.u_0)$.

This completes the proof. \square

Definition 3.7 *Suppose that M is an ordinal DFA and q is a recursive state in M . Define, for each $n \geq 0$, each prefix $u1 \leq_p u_0 = u_0^n$, and each prefix p of u_0 :*

$$\begin{aligned} \mathcal{P}(q, n, p) &:= \{u_0^n p : p \in \mathcal{L}(q)\} \\ \mathcal{Q}(q, n, u1) &:= \{u_0^n u_0 w : u_0 w \in \mathcal{L}(q)\} \\ \mathcal{P}(q, n) &:= \bigcup_{p <_p u_0} \mathcal{P}(q, n, p) \\ \mathcal{Q}(q, n) &:= \bigcup_{u1 \leq_p u_0} \mathcal{Q}(q, n, u1) \\ \mathcal{R}(q, n, u1) &:= \mathcal{P}(q, n) \cup \mathcal{Q}(q, n, u1) \\ \mathcal{R}(q, n) &:= \bigcup_{u1 \leq_p u_0} \mathcal{R}(q, n, u1). \end{aligned}$$

Note that $\mathcal{P}(q, n)$, $\mathcal{Q}(q, n)$ and $\mathcal{R}(q, n)$ are finite unions. Also,

$$\mathcal{R}(q, n) = \mathcal{P}(q, n) \cup \mathcal{Q}(q, n).$$

Thus, by Corollary 3.6,

$$\mathcal{L}(q) = \bigcup_{n \geq 0} \mathcal{R}(q, n).$$

Proposition 3.8 *Suppose that M is an ordinal DFA and q is a recursive state in M . If $0 \leq n < m$, and if $v \in \mathcal{R}(q, n)$ and $w \in \mathcal{R}(q, m)$, then $v <_\ell w$.*

Proof. There are several cases. First, suppose that $v \in \mathcal{P}(q, n)$. If $w \in \mathcal{P}(q, m)$, then either

$$v = u_0^n p,$$

for some $n \geq 0$ and some prefix p of u_0 which belongs to $\mathcal{L}(q)$, and

$$w = u_0^n u_0^{m-n} p',$$

for some $p' \leq_p u_0$ which belong to $\mathcal{L}(q)$. But then $v <_p w$.

If $w \in \mathcal{Q}(q, m)$, then

$$w = u_0^n u_0^{m-n} u'0w',$$

so that again $v <_p w$.

Suppose now that $v \in \mathcal{Q}(q, n)$. If $w \in \mathcal{P}(q, m) \cup \mathcal{Q}(q, m)$, it is easy to see that $v <_s w$. \square

Corollary 3.9 *Suppose that M is an ordinal DFA and q is a recursive state in M . Then $(\mathcal{L}(q), <_\ell)$ is the ordered sum*

$$(\mathcal{L}(q), <_\ell) = (\mathcal{R}(q, 0), <_\ell) + \dots + (\mathcal{R}(q, n), <_\ell) + \dots$$

Proof. By Corollary 3.6 and Proposition 3.8. \square

Let q be a recursive state in an ordinal DFA, and suppose $u1 \leq_p u_0 = u_0^q$. For a fixed $n \geq 0$, we consider the order-type of $\mathcal{R}(q, n, u1) = \mathcal{P}(q, n) \cup \mathcal{Q}(q, n, u1)$. Note that if p and $u1$ are prefixes of u_0 , either p is prefix of u or $u1$ is a prefix of p .

We will find an upper bound for the order-type of $(\mathcal{R}(q, n), <_\ell)$. Using the notation of Definition 3.7, for $u1 \leq_p u_0$, define

$$\begin{aligned} A &= \{u_0^n p : p \leq_p u \text{ \& } p \in \mathcal{L}(q)\} \\ B &= \{u_0^n p : u1 \leq_p p \leq_p u_0 \text{ \& } p \in \mathcal{L}(q)\}. \end{aligned}$$

Proposition 3.10 *$(\mathcal{R}(q, n, u1), <_\ell)$ is the ordered sum*

$$(\mathcal{R}(q, n, u1), <_\ell) = (A, <_\ell) + (\mathcal{L}(q, u_0^n u_0), <_\ell) + (B, <_\ell), \quad (2)$$

so that the order-type of $(\mathcal{R}(q, n, u1), <_\ell)$ is

$$k + \alpha + k', \quad (3)$$

where k is the number of elements in A , and k' is the number of elements in B , and α is the order-type $(\mathcal{L}(q, u_0), <_\ell)$.

Proof. Suppose that $w \in \mathcal{Q}(q, n, u1)$. If $v = u_0^n p \in A$ then $v <_p w$. Indeed, $w = u_0^n u0w'$, for some $w' \in \mathcal{L}(q.u0)$. But since $p \leq_p u$, $v <_p w$. Similarly, if $v \in B$, $w <_s v$. This proves (2).

The order-types of $(\mathcal{L}(q.u0), <_\ell)$ and $(\mathcal{L}(q.u_0^n u0), <_\ell)$ are the same, since $q.u_0^n = q$. We have proved (3). \square

Since $\mathcal{R}(q, n)$ is the (non disjoint) union of the sets $\mathcal{R}(q, n, u1)$, for $u1 \leq_p u_0$, we have the following result.

Corollary 3.11 *For a recursive state q , the order-type of $(\mathcal{R}(q, n), <_\ell)$ is bounded above by a finite sum $\beta_1 + \dots + \beta_m$, where for $i = 1, \dots, m$, $\beta_i = k_i + \alpha_i + k'_i$, with $0 \leq k_i, k'_i < \omega$ and α_i is the order-type of $(\mathcal{L}(q.u0), <_\ell)$, for some prefix $u1$ of u_0 .*

For later use, we point out the following consequence of Corollary 3.11 and Corollary 3.9.

Corollary 3.12 *Let q be a recursive state in an ordinal DFA. Suppose that for each prefix $u1$ of u_0 , the order-type of $(\mathcal{L}(q.u0), <_\ell)$ is less than ω^h , for a positive integer h . Then the order-type of $\mathcal{R}(q, n)$ is also less than ω^h , and the order-type of $(\mathcal{L}(q), <_\ell)$ is at most ω^h .*

Proof. The first statement follows from Corollary 3.11 and the fact that ordinals less than ω^h are closed under finite sums. The second follows from Corollary 3.9. \square

The next definition adopts a similar notion for context-free grammars from [BE10].

Definition 3.13 *Suppose M is any DFA. For any states q, q' , define*

$$q' \preceq q \iff q.v = q',$$

for some word v . Define $[q] = \{q' : q \preceq q' \text{ \& } q' \preceq q\}$.

Two states q, q' are **equivalent** if $q \preceq q'$ and $q' \preceq q$, i.e., they are in the same strong component. The preorder relation $q \preceq q'$ determines a partial ordering on the equivalence classes $[q]$: $[q'] \leq [q]$ if $q' \preceq q$.

Lemma 3.14 *Suppose $[q'] \leq [q]$. Then if M is an ordinal DFA, the order-type of $(\mathcal{L}(q'), <_\ell)$ is at most that of $(\mathcal{L}(q), <_\ell)$.*

Proof. Let v be a word such that $q.v = q'$. Then, for any word $u \in \mathcal{L}(q')$, the word vu belongs to $\mathcal{L}(q)$. Thus

$$u \mapsto vu$$

is an order-preserving map $\mathcal{L}(q') \rightarrow \mathcal{L}(q)$. \square

Definition 3.15 Suppose M is a DFA and q is a state in M . The **height** of q is the number of equivalence classes $[q']$ such that $[q'] < [q]$.

Corollary 3.16 Suppose M is an ordinal DFA. If $q' \in [q]$, the order-types of $(\mathcal{L}(q), <_\ell)$ and $(\mathcal{L}(q'), <_\ell)$ are the same. If q, q' have the same height and $q' \preceq q$, then $q \preceq q'$.

Proof of the last claim. If there is no path $q' \rightsquigarrow q$, then $[q'] < [q]$, so that the height of q is greater than that of q' . \square

Remark 3.17 In a trim DFA, if there is a sink state, there is a unique one, and its height is zero. Conversely, if q is a state of height zero and $(\mathcal{L}(q), <_\ell)$ is well-ordered, then q is a sink state. Otherwise, since both $q.0$ and $q.1$ are in the strong component of q , this contradicts the Main Lemma.

Theorem 3.1 Suppose that M is an ordinal DFA. If q is a state of height h , then the order-type of $(\mathcal{L}(q), <_\ell)$ is at most ω^h .

Proof. We use induction on h .

When $h = 0$, q must be a sink state, by the previous remark. Thus, the order-type of $(\mathcal{L}(q), <_\ell)$ is 0, and $0 < \omega^0 = 1$.

Assume $h = 1$ and q is not recursive. Then both $q.0$ and $q.1$ are the sink. If $q \in F$, the order-type of $(\mathcal{L}(q), <_\ell)$ is 1; if q is not in F , $\mathcal{L}(q) = \emptyset$, showing q is a sink, contradicting the assumption that M is trim.

Assume $h = 1$ and q is recursive. Then each exit edge from the strong component of q labeled either 0 or 1 has the sink as target. There must be some final states in the strong component of q , or else q itself is a sink. Say there are $k > 0$ prefixes of u_0 in F . Since, in this case,

$$(\mathcal{L}(q), <_\ell) = (\mathcal{P}(q, 0), <_\ell) + (\mathcal{P}(q, 1), <_\ell) + \dots$$

we see that the order-type of $\mathcal{L}(q), <_\ell$ is

$$k + k + \dots = \omega.$$

To complete the induction, assume $h > 1$ and suppose that if a state has height less than h , then the order-type of its language is at most $\omega^{h'}$, for some nonnegative integer $h' < h$. If q has height h , either it is recursive, or not. If not, the order-type of q is at most $1 + \alpha_0 + \alpha_1$, where α_i , $i = 0, 1$, is the order-type of $(\mathcal{L}(q.i), <_\ell)$. Since $q.i$ has height less than h , the order-type of $(\mathcal{L}(q), <_\ell)$ is at most $1 + \omega^{h-1} \times 2 < \omega^h$.

If state q has height h and q is recursive, then by Corollary 3.12, the order-type of $(\mathcal{L}(q), <_\ell)$ is at most

$$\begin{aligned} \omega^{h-1} + \omega^{h-1} + \dots + \omega^{h-1} + \dots &= \omega^{h-1} \times \omega \\ &= \omega^h. \quad \square \end{aligned}$$

As a consequence of Theorem 3.1 and Corollary 3.4, we obtain another proof of the following result.

Corollary 3.18 *An ordinal α is regular if and only if $\alpha < \omega^\omega$.*

Proof. By Corollary 3.4, we need prove only that any regular ordinal is less than ω^ω . If α is regular, there is an ordinal DFA M such that α is the order-type of $(\mathcal{L}(M), <_\ell)$. By Theorem 3.1, if M has n states, the order-type of $(\mathcal{L}(M), <_\ell)$ is at most ω^n . \square

4 Summary

Aside from an alternative proof of the result in Corollary 3.18, we have found a structural characterization of ordinal DFAs in Corollary 2.14 and an $O(n^2)$ -algorithm to identify them. It would be interesting to find a structural characterization of those DFAs M such that $(\mathcal{L}(M), <_\ell)$ is

- dense, or
- scattered.

References

- [Ada74] J. Adamek. Free algebras and automata realizations in the language of categories. *Comment. Math. Univ. Carolinae*, 15(1974), 589–602.
- [BC01] S.L. Bloom and C. Choffrut. Long words: the theory of concatenation and ω -power. *Theoretical Computer Science*, 259(2001), 533–548.
- [BE10a] S.L. Bloom and Z. Ésik. Algebraic Linear Orderings. to appear.
- [BE10] S.L. Bloom and Z. Ésik. Algebraic Ordinals. to appear in *Fundamenta Informaticae*.
- [BE07] S.L. Bloom and Z. Ésik. Regular and algebraic words and ordinals. In: *CALCO 2007, Bergen*, LNCS 4624, Springer, 2007, 1–15.

- [BE10] S.L. Bloom and Z. Ésik. A Mezei-Wright theorem for categorical algebras. *Theoretical Computer Science* 411 (2010) 341–359.
- [CLRS] T. H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein. *Introduction to Algorithms*, Third Edition. The MIT Press. Cambridge, MA., 2009.
- [Cour78a] B. Courcelle. Frontiers of infinite trees. *RAIRO Theoretical Informatics and Applications*, 12(1978), 319–337.
- [Cour78] B. Courcelle. A representation of trees by languages, *Theoretical Computer Science*, 6 (1978), 255–279 and 7(1978), 25–55.
- [Cour83] B. Courcelle. Fundamental properties of infinite trees. *Theoretical Computer Science*, 25(1983), 95–169.
- [Del04] Ch. Delhommé. Automaticity of ordinals and of homogeneous graphs. *C. R. Math. Acad. Sci. Paris* 339(2004), no. 1, 5–10. (in French)
- [Heil80] S. Heilbrunner. An algorithm for the solution of fixed-point equations for infinite words. *RAIRO Theoretical Informatics and Applications*, 14(1980), 131–141.
- [KRS03] B. Khoussainov, S. Rubin and F. Stephan. On automatic partial orders. *Proceedings of Eighteenth IEEE Symposium on Logic in Computer Science, LICS*, 168-177, 2003.
- [Roit90] Judith Roitman. *Introduction to Modern Set Theory*. Wiley, 1990.
- [Ros82] J.B. Rosenstein. *Linear Orderings*. Academic Press, New York, 1982.
- [Wand79] M. Wand. Fixed point constructions in order-enriched categories. *Theoretical Computer Science*, 8(1979), 13–30.